Diagnosis of DCIS (stage 0 breast cancer), IDC (stage I cancer) and colon cancer using methods of the invention.

This document serves to provide details of experiments which have been conducted by their inventors in their laboratory in which differentially expressed transcripts have been identified in samples from DCIS, IDC and colon cancer patients relative to healthy patients and these transcripts have been used as probes to predict whether samples are derived from patients with the disease in question or from patients without that disease. The samples which have been examined are peripheral blood samples. The experiments which have been conducted are described in the following paragraphs.

Experimental details

Whole blood from the arms of 48 females were collected in PAX tubes at Ullevål University Hospital Norway. The mean age of females was 58.16 with an age range of 51- 82. These included 20 females diagnosed with breast cancer, twenty control females (termed as healthy) who had suspected first mammograms, but were later diagnosed as not having breast cancer, and eight females having another form of cancer, colon cancer. Among those having breast cancer, ten females had stage 0 cancer (Ductal Carcinoma In Situ, DCIS in which the cancer is confined within the milk ducts) with a tumour size ranging between 4 mm and 50 mm and ten females had stage I cancer (Invasive Ductal carcinoma, IDC) with a tumour size ranging between 2 mm and 19 mm. Details of the breast cancer patients used in the study are provided in Table 1.

Total RNA was extracted from the blood samples and subjected to reverse transcription to yield first strand cDNA from which DIG-cRNA probes were prepared by amplification. The probes were hybridized to high density arrays (Applied Biosystems Human Genome Survey Microarray v2.0) containing 32,878 oligonucleotide probes. The amount of labelled probes binding to the immobilized oligonucleotides was assessed and quantified using the Applied Biosystems 1700 Chemiluminescent Microarray Analyzer.

The generated expression data were then processed. The data were normalized to take account of the differences in the probe intensities resulting from the experimental conditions. The data set was then analyzed to identify differentially expressed informative probes for different sample groups. Informative probes were identified for the following set ups:

Set up 1: DCIS&IDC vs Healthy samples.
Set up 2: DCIS vs Healthy samples.
Set up 3: IDC vs Healthy samples.
Set up 4: Colon cancer vs Healthy samples.

Several genes were found to be informative in the different set ups. Table 2 provides a matrix of informative genes in each set up and the number of informative genes that

were found overlapping among the different set ups. Thus for example in the case of set up 1, 197 genes were identified as informative and reflect differential expression between DCIS&IDC samples vs healthy samples. Of these 197, 133 were also found to be informative in set up 2, i.e. for discrimination between DCIS and healthy samples.

To illustrate the utility of these informative probes diagnostically, classification models were generated. The classification models were able to classify individuals from different classes into distinct group e.g. DCIS from healthy samples (data not shown). The ability of the general model to correctly diagnose samples was determined by cross-validation in which the step of generating the classification model was performed by omitting the data of a single sample from the data used in that modelling process. This process was repeated for each sample to obtain information on the prediction accuracy of the calibration model. The prediction result of the classification model for the 4 set ups is shown in Figures 1A-1D. In the 4 prediction plots, the diseased samples appear on the x axis at +1 and the healthy samples appear at -1. The y axis represents the predicted class membership. During prediction, if the predicted class is correct, the diseased samples should fall above zero and the healthy samples should fall below zero. Correct prediction was achieved for all samples in set ups 2-4. In set up 1 only a small number of samples were incorrectly predicted.

This illustrates that cancer samples can be readily discriminated from normal samples using the identified probes (Figures 1A-1D). Figure 1B shows that probes which are informative for DCIS relative to healthy patients can be identified. Since, in the case of DCIS, the cancer is confined within milk ducts, these results show that blood that does not contain cancer cells and the cells of which have not been in contact with the diseased area, exhibit characteristic changes in their gene expression pattern and this pattern can be used diagnostically for early detection of breast cancers.

The ability of the built-in crossvalidated models to predict the class of a sample group that was not included during model-building process (test set prediction) is presented in Figures 2A - 2D. In these prediction results, samples below 0 are classified as healthy and samples above 0 are classified in the diseased group.

The set up 1 probes which are able to distinguish between DCIS&IDC and healthy samples were used to predict the class of colon cancer samples. It will be seen from the results in Figure 2A, that the cross-validated model correctly predicted the class of 6/8 colon cancer samples as non-breast cancer, i.e. below 0. This result shows that the identified probes for breast cancer are specific and can efficiently discriminate breast cancer from other forms of malignancies.

The results in Figure 2B show that probes identified as being informative for discriminating between DCIS and healthy samples can also be used to predict the category of IDC patients as cancer samples. In the model 8/10 IDC samples were correctly predicted.

In case of set up 3, the built-in cross-validated model based on informative probes which discriminated between IDC and healthy samples correctly predicted the class of 8/10 DCIS samples (Figure 2C). This illustrates that the altered expression in the IDC

samples reflects at least some alterations also seen in DCIS samples. Thus this provides evidence that the cells being examined are cells which are not cancer cells and have not been in contact with the disease area.

In case of set up 4, the built-in model correctly predicted the class of 19/20 breast cancer samples as non-colon cancer (Figure 2D). The result shows that these samples were effectively predicted as "healthy" samples, i.e. the informative probes are able to distinguish between different types of cancers.

Hence the results of this study show that:

- Probes can be identified from blood cells for the diagnosis of early stage breast cancer when the blood cells have not contacted the diseased area.
- Probes can be identified which discriminate between different types of cancer.

Table. 1: Details of the breast cancer patients used in the study

Stage 0 - DCIS

Sample number	Age	Breast cancer subtype	Histology
1	70	DCIS	25mm
2	64	DCIS	4mm
3	62	DCIS	5mm
4	65	DCIS	30mm
5	59	DCIS	30mm
6	65	DCIS	20mm
7	51	DCIS	12mm
8	na	DCIS	26mm
9	na	DCIS	2mm
10	na	DCIS	50mm

Stage I - IDC

Sample number	Age	Breast cancer subtype	Histology
11	58	IDC	4mm
12	69	IDC	17mm
13	51	IDC	15mm
14	57	IDC	8mm
15	57	IDC	12mm
16	68	IDC	19mm
17	64	IDC	9mm
18	53	IDC	<20mm
19	60	IDC	7mm
20	68	IDC	15mm

Table 2. Matrix of differentially expressed informative genes in the different setup. H: healthy; CC: colon cancer. Breast cancer samples includes both stage 0 (DCIS) and stage I (IDC) samples.

	Set up 1 (breast cancer/H)	Set up 2 (DCIS/H)	Set up 3 (IDC/ H)	Set up 4 (CC/ H)
Set up 1 (breast cancer/H)	197	133	106	17
Set up 1 (DCIS/H)		647	88	37
Set up 3 (IDC/ H)			494	36
Set up 4 (CC/ H)				735

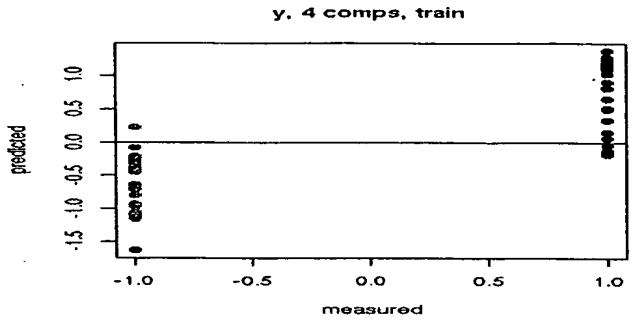
Figure 1. Prediction plot based on informative genes identified in the four different set ups. 1**A:** set up 1, breast cancer (both DCIS and IDC versus healthy samples. 1**B:** set up 2, DCIS versus healthy samples. 1**C:** set up 3, IDC versus healthy samples. 1**D:** set up 4, colon cancer versus healthy samples. Healthy samples appear on the x-axis at -1 and diseased samples appear at +1.

Figure 2. Prediction results. 2**A:** Prediction of colon cancer samples by the model based on breast cancer and healthy samples. 2**B:** Prediction of IDC samples by model based on DCIS and healthy samples. 2**C:** Prediction of DCIS samples by model based on IDC and healthy samples. 2**D:** Prediction of breast cancer samples by model based on colon cancer and healthy samples.

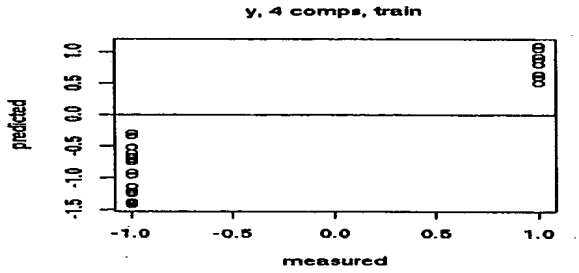
FIGURE 1

Prediction plots (Breast cancer study)

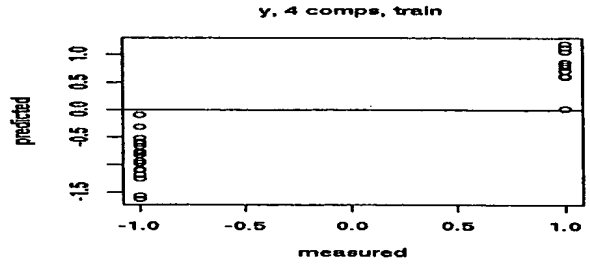
A **Set up 1**



B **Set up 2**



C **Set up 3**



D **Set up 4**

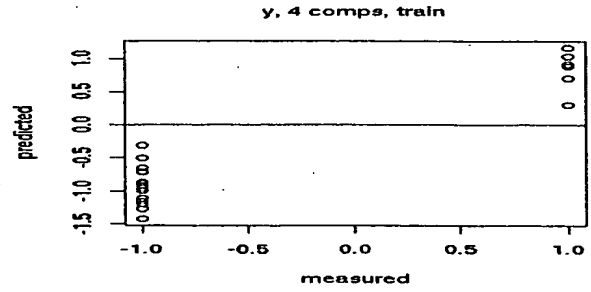
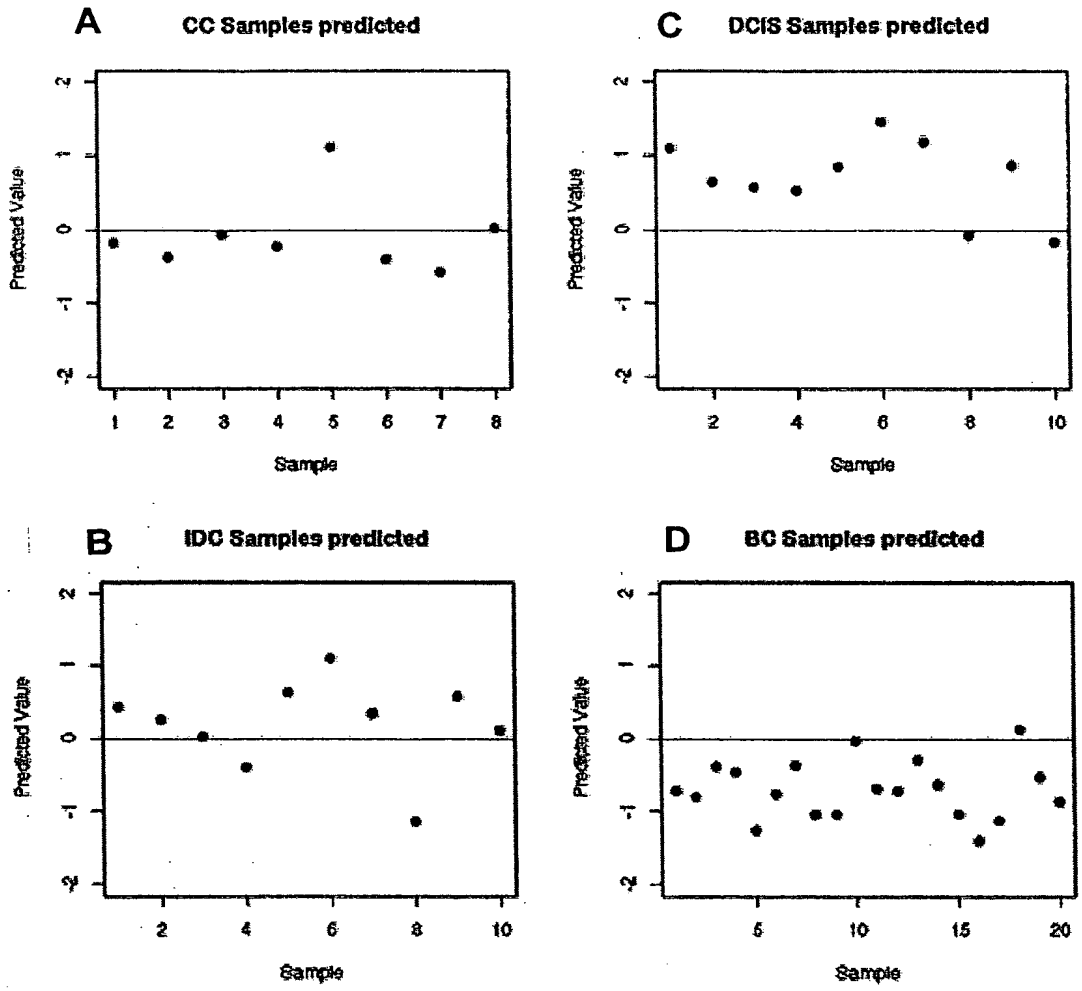


FIGURE 2


A
CC Samples predicted
Predicted Value
Sample
B
IDC Samples predicted
Predicted Value
Sample
C
DCIS Samples predicted
Predicted Value
Sample
D
BC Samples predicted
Predicted Value
Sample